ARI Research Note 88-93

Empirical Demonstration of Isoperformance Methodlogy Preparatory of an Interactive Expert Computerized Decision Aid

Robert S. Kennedy, Marshall B. Jones,
and Dennis R. Baltzley
Essex Corporation

for

Contracting Officer's Representative
Donald B. Headley

Manned Systems Group
John F. Hayes, Chief

Systems Research Laboratory
Robin L. Keesee, Director

November 1988

Accesion For		
NTIS CRA&I	☑	
DTIC TAB	☐	
Unannounced	☐	
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

United States Army
Research Institute for the Behavioral and Social Sciences

AD-A202 439

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

DTIC SELECTED DEC 2 0 1988 D

DTIC FILE COPY

1a. REPORT SECURITY CLASSIFICATION: - -	1b. RESTRICTIVE MARKINGS: - -
2a. SECURITY CLASSIFICATION AUTHORITY: - -	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution is unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE: - -	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): - -	5. MONITORING ORGANIZATION REPORT NUMBER(S): ARI Research Note 88-93
6a. NAME OF PERFORMING ORGANIZATION: Essex Corporation; 6b. OFFICE SYMBOL (If applicable): 1D B06	7a. NAME OF MONITORING ORGANIZATION: U.S. Army Research Institute for the Behavioral and Social Sciences
6c. ADDRESS (City, State, and ZIP Code): 1040 Woodcock Road, Suite 227, Orlando, Florida 32803	7b. ADDRESS (City, State, and ZIP Code): 5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: see 7a; 8b. OFFICE SYMBOL (If applicable): PERI-SM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: MDA903-87-C-0603
8c. ADDRESS (City, State, and ZIP Code): see 7b	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. 6.55.02.A; PROJECT NO. 2Q6655 02M770; TASK NO. (SBIR); WORK UNIT ACCESSION NO. - -

11. TITLE (Include Security Classification): Empirical Demonstration of Isoperformance Methodology Preparatory to Development of an Interactive Expert Computerized Decision Aid

12. PERSONAL AUTHOR(S): Robert S. Kennedy, Marshall B. Jones, and Dennis R. Baltzley

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Report	FROM 87-8 TO 88-2	November 1988	41

16. SUPPLEMENTARY NOTATION: Donald B. Headley, contracting officer's representative. Manpower and Personnel Integration; Manpower Personnel Training

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Procurement and Acquisition, Strategies, MPT Issues, Human Factors Engineering, Systems Design, MANPRINT. (SDW) Job Performance, Isoperformance,

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This research note reports on the increasing need within the Department of Defense to improve systems performance through better integration of men and women into military systems. The underlying theme of the Research Note is that once the level of operational performance is settled upon, trade-offs among available resources can and should be made. Formal methodologies are proposed which permit mixtures of training, equipment, and individual differences to be traded off to arrive at a prescribed level of systems performance. This approach is called "isoperformance". Keywords.

Reproduced From Best Available Copy

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION: Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL: Donald B. Headley	22b. TELEPHONE (Include Area Code): 202/274-8917	22c. OFFICE SYMBOL: PERI-SM

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE: UNCLASSIFIED

i

20000920173

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract for the Department of the Army

Essex Corporation

Technical review by

Donald B. Headley
Uldi Shvern

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Informational Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not to be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

EMPIRICAL DEMONSTRATION OF ISOPERFORMANCE METHODOLOGY PREPARATORY TO DEVELOPMENT OF AN INTERACTIVE EXPERT COMPUTERIZED DECISION AID

CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

EMPIRICAL DEMONSTRATION OF ISOPERFORMANCE METHODOLOGY PREPARATORY TO DEVELOPMENT OF AN INTERACTIVE EXPERT COMPUTERIZED DECISION AID

INTRODUCTION

Overview

Human performance in complex systems is a function of human-machine interaction. Within systems engineering, such interaction is the focus of attention by design engineers on the one hand, and behavioral scientists on the other. The behavioral scientists are variously listed as being in human factors, human factors engineering, engineering psychology, and (less often) applied experimental psychology. Although the first text devoted to the subject of "men and machines" was labeled "Human Engineering," the authors (Chapanis, Garner, Morgan, & Sanford, 1947) preferred the "more accurate and definitely more cumbersome term...psychophysical systems research" (p. 5) and they called their facility the Systems Research Laboratory.

In the early work these pioneers acknowledged their lineage in experimental psychology and offered that "personnel" and "educational" are other fields related to systems research, but "personnel selection has developed to such an extent that it is now a relatively complete and independent branch of psychology...[and]...we in the Systems Research Laboratory...are not primarily interested in this aspect of the total problem" (p. 10). Since at least that time the fields of systems, training and selection have remained largely independent. Their methods are different. Personnel emphasizes the use of correlational analyses. Education and training employ repeated measures. In systems research and engineering psychology, the focus is often on point-and-error range estimates of human lawful relationships (transfer functions) from independent variable manipulations.

Typically, in systems research work, it has been taught that, as human factors practitioners, it is our role to gather human input/output data (transfer functions) of man with his equipment (or physical and environmental stimuli). These data would then be used to generate standards and specifications which could be used by design engineers and this would thereby improve systems performance.

It was believed that design engineers were eagerly awaiting these data to incorporate into new systems which would permit efficient allocation of functions between man and machines. This goal while lofty, was naive, and one of the intentions of this report is to call attention to a technique whose goal is to improve decision-making human engineering in systems research and which embraces and uses as a theme the notion of "trade-off technology." This approach deals with total or operational systems performance and focuses on the premise that differing combinations of individual differences, training, and equipment variables can lead to the same desired outcome. It is called isoperformance (iso meaning same) and is a conceptual approach to systems research in human engineering. The focus of isoperformance is that the same level of performance can be attained by different combinations of personnel, training, and equipment. The goal is that, once these combinations have been determined, choices among them can be made in terms of maximum feasibilities or minimum costs. The program takes into account equipment and systems, personnel, and training research. It leaves an audit trail of the decision process.

The report is divided into five sections which form an integrated look at isoperformance. The first section outlines the literature from which isoperformance was conceived and gained its foundation. The second provides a straightforward empirical test which shows that isoperformance does indeed work and provides detailed analyses and descriptions of how it works. The objective was to show a "proof-of-concept," to provide empirical support for the isoperformance approach, and to demonstrate how it may be applied to a real world situation. Section three deals with the key issue of available alternatives where incomplete data exist and provides suggestions for action. Section four provides the range of uses for isoperformance and the last section highlights systems, themes, and directions in which isoperformance may head.

The idea for the isoperformance methodology emerged from the authors' previous experiences with the experimental conduct of flight simulation studies, and the use of multivariate analyses of the data (e.g., Lintern, Nelson, Sheppard, Westra, & Kennedy, 1981). These studies followed a review of human factors engineering experiments (Simon, 1976) where it was concluded that the methods most commonly used were often misapplied or inadequate for obtaining the desired information. In Simon's analysis, a quantitative evaluation of the quality of the data produced in human factors engineering experiments and the methods employed to obtain these data were presented. The data were reported as distribution and "proportions-of-variance-accounted-for" by experimental factors in 239 experiments. His discovery was that equipment factors accounted for less variance than subject and other factors like practice, at least when subject and practice factors were seriously interpreted. But as the number of factors in an experiment was increased, increasing proportions of variances became attributable to equipment features.

The authors of this report have been associated with experiments at the Navy's Visual Technology Research Simulator for several years and these efforts have followed Simon's holistic methodologies and provide general support for this projection. In these studies, although the amount of variance accounted for by equipment features is not a large proportion of total experimental variance, it should be noted that the worst combination of equipment features never results in an "unflyable" simulation and so that dimension has a range restriction (Westra & Lintern, 1985). On the other hand, the subject variables (usually aviators) and training variables (often experienced pilots) are also restricted in range, yet they appear to account for larger proportions of variance. In fact, in one experiment in which 10 simulator equipment factors, including major cost variables like simulator motion and field of view were tested, all of the equipment factors combined accounted for less variance than the reliable pilot differences of highly experienced fleet pilots (Westra, Simon, Collyer, & Chambers, 1982).

The studies from the Navy's Visual Technology Research Simulation program (Lintern et al., 1981) contained encouraging results for a conceptual model like the isoperformance notion proposed here. In experimental studies of the effects of performance and equipment, including individual differences, one emerges from the analyses with a breakdown of the total variance attributable to each of the main effects "equipment," "training," "aptitude," and some interactions of these (cf. Kennedy, Berbaum, Collyer, May, & Dunlap, 1983). The general finding in analyses of studies of this sort is that the individual differences or aptitude variables account for a substantial proportion of the total explained variance, and more than either practice or equipment variations (Lintern & Kennedy, 1984; Westra & Lintern, 1985; Westra et al., 1982). Furthermore, as a rule, practice accounts for more than equipment (Lintern et al., 1981). This finding permitted a potentially useful inference about the importance of the three major components in the determination of performance at the end of appreciable lengths of practice. However, the generality of this notion to the system research literature in general was unknown. Missing, therefore, was an explicit understanding of the trade-offs among the three major components relative to producing a given level of performance.

A meta-analysis (Green & Hall, 1984) of the systems research and human factors engineering literature was therefore conducted which compared these three types of variables (Jones, Kennedy, Turnage, Kuntz, & Jones, 1986). The analysis went beyond the time-frame used in Simon's review and sought to determine whether the human factors studies (Lintern et al., 1981) in the Navy simulator would generalize to the scientific literature in human factors engineering. Green and Hall (1984) list several methods ranging from simple (e.g., box-score tally of the direction of effect) to more sophisticated, descriptive (e.g., size of the effect or d prime [Swets, Tanner, & Birdsall, 1961]) and more inferential (e.g., eta squared, omega squared [Hays, 1977]).

It was decided to follow an inferential (omega squared - Hays, 1977) approach and a quantitative analysis was settled upon for those studies identified as suitable for such calculations. This calculation is a normalized measure of relationship which permits quantitative comparison between experiments with widely differing characteristics in sample size, training methods, and equipment options. To this end, studies in the human factors engineering literature were identified which examined at least two of the following variables together: practice or training, individual differences, and equipment features. The review included a computerized search at the University of Central Florida through the NASA-Southern Technology Applications Center (STAC) data base. The National Technical Information Service (NTIS), NASA, and human factors literature were reviewed. A list of key words to be used in the computer literature search was generated. Venn diagrams were used to structure the search and otherwise filter out the literature that was not of interest. For example, over 11,000 articles were catalogued under the subject heading "Human Factors Engineering." However, the combination of "Human Factors Engineering" and "Training/Learning" yielded 153 articles (30 of which were classified). Combining terms in this manner made the number of citations to review a much more manageable figure. Of over 10,000 titles searched, 276 involved experimental studies of training and performance as a function of equipment variations; 68 involved an analysis of variance; 30 reported ANOVA data; but only 10 permitted sufficient detail for calculation of omega squared. This final yield was a miniscule .1% of the original number, an important and somewhat sobering commentary on the raw material that serves as the technological data base for systems research and human factors engineering.

Moreover, although the meta-analysis of the 10 studies for which sufficient data were available was revealing, it was also disappointing. It showed that there is no difficulty in the calculation of omega squared if the experimental outcomes are fully reported and the designs adequately conceptualized. Unfortunately 10 studies are too few and the data turned out to be too irregular to permit sufficient generalizations about trends in these studies. Certainly there is insufficient regularity in published studies to implement in an isoperformance model. For example, three of the five studies with high omega-squared values for subjects involved no equipment variation. Thus, the absence of an equipment variation did not explain the high value of omega squared. A similar situation prevailed among the four studies with low values of omega squared: two involved several important equipment variations but the other two did not. Therefore, it was impossible to integrate the findings of these reports even when they contained the necessary ANOVA information because of the multiplicity and noncomparability of fixed-effect measures. This result carries the clear implication that a meta-analysis of the existing literature will <u>not</u> suffice to implement the isoperformance or any other empirical trade-off approach. This is not to say that there could not be valuable lessons learned from the literature, but that the literature in its present form will not permit definitive answers. It should be noted that recently in a formal meta-analysis of more than 12 studies of simulator equipment features, (Jones, Kennedy, Baltzley, & Westra, in preparation) it has been found that on the average twice as much of the reliable (main effect) variance is due to subjects as to training and equipment variance combined.

There are several options available. One, technologists can familiarize themselves with the literature and then they can be heavily constrained to make estimates about relationships. This possibility has been explored somewhat in our USAF interactive computer program (Jones, Kennedy, Kuntz, & Baltzley, 1987). Alternatively, if extrapolations from the existing literature to real-world situations are to be made, they are going to have to be exemplified by formal experiments carried out for the purpose and implemented under an innovative technical framework. Such a framework can be proposed with a developmental effort into isoperformance and in greater detail with experimental exemplification of the framework and application (i.e., validation) in a real-world situation. While other methods for conducting human factors research exist, it is believed most, if not all, fall short of total system consideration.

Several methodologies now exist for the implementation of psychophysical system research and engineering design criteria and standards, and modern manuals and handbooks are available for guidance (viz., Boff, 1984; Department of Defense, 1981; Malone, Shenk, & Moroney, 1976; Morgan, Cook, Chapanis, & Lund, 1963; Perkins, Binel, & Avery, 1983; Woodson, 1955). Human performance models for man-machine systems evaluation are available (cf. Pew, Baron, Feehrer, & Miller, 1977, for a review). Over the past 20 years, much of the improvement in these systems approaches has been in an emphasis on test and evaluation rather than on design (Kearns, 1982). "Reverse engineering" (Marcus & Kaplan, 1984) is an attempt at feeding back into systems design the conclusions that most affect human factors manpower and training considerations. The application of reverse engineering represents a direct recognition that human factors, manpower, personnel, and training are critically important inputs in the weapons acquisition process.

Similarly, the Manpower and Personnel Integration (MANPRINT) initiative makes the following domains imperative in the materiel acquisition process: human factors engineering; manpower/personnel/training (MPT); systems safety research, and health hazard assessments (cf. U.S. General Accounting Office, 1985 for a bibliography of relevant studies within the three military services). One important MANPRINT contribution to research and development for materiel acquisition is the origination of generic analytic tools for answering important allocation questions such as can soldiers operate equipment effectively, how do complex man-machine systems work, and how much and what kind of training is needed? A generic analytic tool, Hardware versus Manpower (HARDMAN) (Mannle, Guptill, & Risser, 1985) provides a baseline comparison methodology and uses operational concepts to predict MPT needs. This type of analysis provides information about required sustainment costs, training costs, and projects how many people will be needed to service and operate systems in the field. Additionally, many other generic design modeling systems are currently available such as HOS (Human Operator Simulator) and HOS-IV (Harris, Iavecchia, Ross, & Shaffer, 1987) and SAINT/MicroSAINT (Laughery, Drews, Archer, & Kramme, 1986) to develop operational concepts in laboratories before any money is spent to build weapon systems.

Despite MANPRINT and other attempts to use human factors engineering and systems analysis to help man-machine systems reach maximum performance within specified constraints, it is believed that inadequate attention appears to be paid to individual differences and training as related to human factors engineering design. Moreover, neither of these are well incorporated into military standards in any formal way. Therefore, they are largely ignored in the design of equipment. A known exception is the leverage that can be applied by modelling anthropometric differences between members of a user population (cf. Bittner & Moroney, 1984, 1985, for a description of this approach). Examples of individual differences and training and how they may impact on suitable design of systems now follows.

Individual Differences

These differences include all of the many identifiable variations in people from sensory sensitivities and anthropometric variances to mental capabilities. Military personnel are selected along many dimensions of individual differences. For example, anyone classified below Category 4 on the Armed Forces Qualifications Test (AFQT) (Maier & Grafton, 1980) are not accepted into service. Nevertheless, even with these restrictions in range (Sims & Hiatt, 1981), individual differences among military personnel are great. For example, the distance at which one pilot customarily detects opponent aircraft is sometimes 50-70% better than another, resulting in 2-3 mile advantages in early detection (Jones, 1981, personal communication). This finding has obvious implications for winning in air combat (Ault Committee Report, 1969, Campbell, 1970). Moreover, some pilots who are better at visual detection can even "outsee" the poorer ones when the latter use telescopes (Jones, 1981, personal communication). In this example, if equipment factors were evaluated to determine effects on performance in terms of the amount of accountable variance, one could not adequately assess the question without taking into account the differing performances of the individual pilots.

Cognitive and other mental capabilities also show wide variation (cf. Schoenfeldt, 1982, for a review). There are also substantial individual differences in basic information processing capacities (Rose, 1978). For example, the speed of mental rotation which may be of utility for photointerpretation varies considerably across individuals. A recent study (Hunt, 1984) found that the fastest subject could perform a mental rotation at approximately 2.5 degrees per msec compared to 18.5 degrees for the slowest subject. Men are generally faster at rotation than women, and young adults are generally faster than people in their 30s and beyond (Berg, Hertzog, & Hunt, 1982). This factor could be the source of the gender effect in video game research, motion sickness, or field independence studies. Moreover, among good readers by general population standards, there are substantial variations in the speed of lexical identification. In one study, there was approximately a 25% variation in speed (560 to 700 msec) between the faster and the slower lexical decision makers (Hunt, Davidson, & Larsman, 1981; Palmer, McLeod, Hunt, & Davidson, 1983). People also vary markedly in the number of sentences that they can process while still being able to recall the words. College students show differences of 2 to 5 sentences, and people who show more "verbal aptitude" seem to have markedly longer spans (Daneman, 1983).

While mental competence is apparently bounded by a person's information processing capabilities, there are very large variations in performance within these bounds which may be attributable to differences in problem solving strategy and by knowledge of a content area. For instance, one study explored models of strategy and strategy shifting on a spatial visualization task using high school and adult subjects (Kyllonen, Woltz, & Lohman, 1981). For each of three successive task steps (encoding, construction, and comparison) separate models applied for individual subjects, suggesting that subjects used disparate strategies for solving the same items. Numerous other studies (e.g., Yalow, 1980) provide evidence that neither aptitude nor instructional treatment alone can fully describe learning and performance outcomes. Interactions between them exist and are consistently demonstrated. Instructional supplements can effectively "fill-in" for student weaknesses and reduce differences between high and low ability students. However, such supplements must be used with caution because reducing the difficulty of instructional materials may enhance immediate learning but fail to display any long-term advantages.

At the physical end of the human performance spectrum, muscular strength (Alluisi, 1978, p. 354) also shows sufficiently wide variances such that in tasks which require upper body lifting, one would find that the 95th percentile female could not perform as well as the average male. At the more global end of human capability, team performance in tanks is largely a function of the intelligence of the tank commmander (Wallace, 1982).

In summary, individual differences such as these have obvious implications for human factors engineering design because they can overshadow the effect of equipment modifications. Yet there is no formal mechanism to incorporate them into military standards, nor do any of the manpower management systems deal with them effectively.

Training

Recently, a review of the lawful relationships from the scientific literature related to military training has been completed for DoD (Lane, 1986). The sheer magnitude of the information in the report defies simple explanation. Learning curves vary in their shape. Tasks that are primarily conceptual may show plateaus or large gains with short amounts of practice. Skill acquisition and procedural tasks, however, generally show the "traditional learning curve. The shape of the learning function is such that the most rapid amount of training effect occurs initially and the best description of the overall relationship is that log performance (or practice) is a linear function of log practice (Newell & Rosenbloom, 1981). Thus, ranges of improvement in performance during military training in formal schools can be an order of magnitude of improvement for each epoch of time spent in training (cf. Hagman & Rose, 1983; Lane, 1986; Schendel, Shields, & Katz, 1978, for reviews). Therefore, improvements of as much as 500% are not unusual. It follows that tasks which can only be performed with great difficulty and extreme concentration initially may be performed with far less mental attention after modest amounts of practice. Moreover, the advantages of display aiding (e.g., Smith & Kennedy, 1976) or artificial intelligence may be largely during these initial stages and of far less utility when the learning curve has slowed down. Such a range of improvements can temper any expected change due to equipment factors.

Although some of these findings have been used for decision making in industrial settings they appear not to have found their way into existing manpower management models like the Navy's HARDMAN, the newer Air Force program RAMPART, and MANPRINT. Furthermore, improvements with practice can be compounded by the fact that there are also large individual differences in practice effects. For example, Kennedy, Bittner, Harbeson, and Jones (1982) found that performance improvement on a video game task proceeded at very different rates, and some of those who learned slowly at first eventually outperformed the fast learners if sufficient trials were given. Because of large individual differences in rates of learning, accuracy of prediction suffers when performance data are collected too early. Furthermore, these aptitude by treatment interactions (ATI; Snow, 1980) have shown that the correlation of general ability or aptitude to acquisition rate tends to increase as instruction places increased information processing burdens on learners, and the correlation decreases as instruction is designed to reduce the information processing demands on learners. Equipment features too can interact with ability. Wightman and Lintern (1985) found that the advantages of part-task versus whole-task relationships were different depending on aptitude. A large literature (some of which is reviewed in Harbeson, Bittner, Kennedy, Carter, & Krause, 1983; Lane, 1986) is available showing representative ranges of these relationships.

The problem outlined above is not one which will lessen with time, but rather the converse. It is believed that the problem of function allocation becomes more critical with the growing complexity and sophistication of machine systems. Since the publication of a landmark article by Fitts in 1951, little progress has been made toward the solution of this problem. Fitts proposed what is now informally called the "Fitts list." This two-column list compares one column headed by the word "man" and another column headed by the word "machine." Fitts' recommendation was to compare the functions for which man is superior to machine to the functions for which the machine is superior to man. While rational, this formulation has yielded little progress in the understanding of systems research interactions and tells little about how to determine trade-off allocations of function (Jordan, 1963). The 27-year old comment by Swain and Wohl (1961) is still current: "There is no adequate systematic methodology in existence for allocating functions between man and machine. It is our view this lack is the central problem in human factors engineering today" (p. 1). Considering the survey of the literature cited above, it is believed a systematic methodology can be provided to account for man/machine interface problems and present decision aids to create trade-off alternatives from the human side of the combination, with no loss of operational proficiency. This methodology is called "isoperformance."

Isoperformance Methodology

A cost-effectiveness method may proceed in either of two general ways. The more familiar is to fix costs and maximize effectiveness. One gets, as the popular phrase puts it, "the biggest bang for the buck." The alternate procedure is to fix effectiveness and minimize health, safety, personnel, training, equipment, and manpower costs -- to get "the same bang in the least costly and most expeditious way." This latter approach leads naturally to trade-offs among the cost factors and is the approach taken by isoperformance methodology (Jones et al., 1987).

The heart of this methodology is the isoperformance curve. With respect to aptitude levels and training times such a curve looks like the one given in Figure 1. The Y-axis is aptitude as measured, for example, by the AFQT. The AFQT is a component of the Armed Services Vocational Aptitude Battery (ASVAB) used to define the mental categories on which the overall mental ability of service personnel is reported to Congress (Sims & Hiatt, 1981). The X-axis is training time in weeks. The job might be MOS 95B10, military police. The curve drawn is for 80% proficient. That is, any point on the curve (any of the indicated combinations of aptitude level and training time) will produce soldiers 80% of whom are proficient at the job. Thus, if one has high-aptitude soldiers (for example, mental categories 1 and 2 on the AFQT) 80% proficient can be reached in roughly eight weeks. With lower aptitude soldiers, more training time is needed and for some aptitude levels (mental category 4 on the AFQT, perhaps) no amount of training time up to the maximum considered will suffice to produce soldiers 80% of whom are proficient.

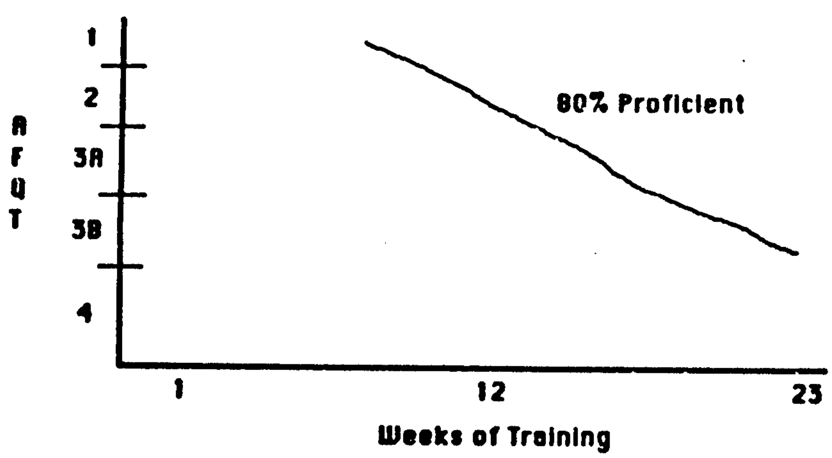

Figure 1. An isoperformance curve for 80% proficient.

Isoperformance curves come in families. A separate and distinct isoperformance curve exists for every level of performance that one specifies. Thus, if one were to specify 50% proficient, for example, one would get a different curve than the one that appears in Figure 1. Note that the second curve (Figure 2) lies to the left and down from the first curve presented. It takes less time to train the same soldiers to the lower level of performance or, in the alternative, for the same amount of training time the lower level of proficiency can be attained with lower aptitude soldiers.

A pair of curves quite similar to the pair in Figure 2 can be obtained in a quite different way. Suppose one were to automate part of the military police job, by providing him/her, perhaps, with computer equipment that was itself easy to use. With the new equipment the job becomes considerably simpler, so that the same objective results can now be achieved by lower aptitude soldiers or with less training time. The situation is depicted in Figure 3. Again there are two curves, but this time the two curves correspond to two equipment variations and both represent the same level of performance. Any point on either curve suffices to produce soldiers 80% of whom are proficient. Using the new equipment the same soldiers can be trained to the same level of performance (80%) proficient) in less time. Or, for a given amount of training time, the same level of performance can be achieved with lower aptitude soldiers.

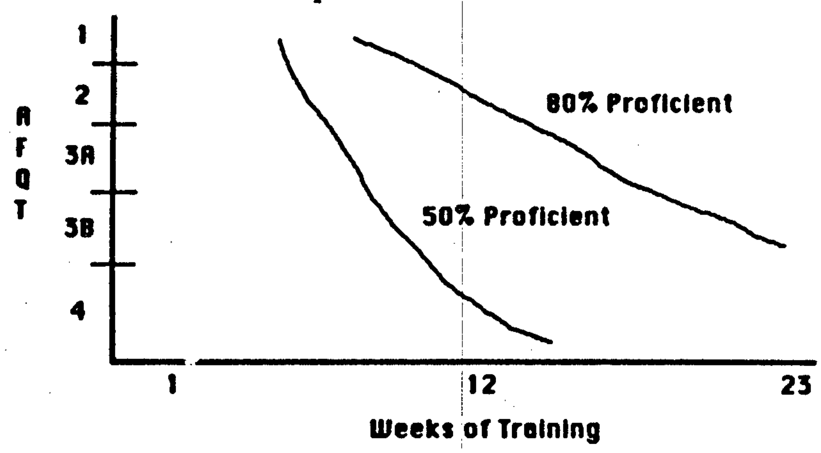

Figure 2. Two isoperformance curves, one for 80% and the other for 50% proficient.

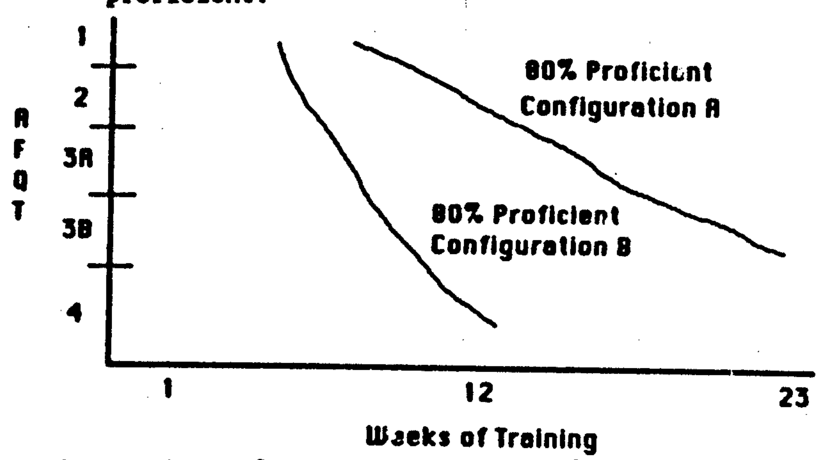

Figure 3. Two isoperformance curves, one for each of two equipment configurations, but both for the same job and the same level of performance.

Isoperformance curves must be evaluated before any conclusion can be reached. Any point on either of the two curves in Figure 3 will produce 80% proficient soldiers -- but which point is best? To answer this question one invokes other cost considerations. Category 1 and 2 soldiers may be in such demand for other jobs that they must be regarded as unavailable. Training times in excess of 12 weeks may be excessively expensive. Figure 4 re presents Figure 3 marked to reflect these two considerations. Since category 1 and 2 soldiers are excluded by reason of unavailability, and category 3 soldiers (or lower) require more than 12 weeks to reach 80% proficient using the original equipment, there is no solution to be obtained using equipment configuration A. The alternative equipment, however, does provide a range of solutions. Any point on the lower curve between the horizontal and vertical bars would be acceptable insofar as personnel availability and training costs are concerned. They might not be equivalent, however, on other counts. It might be, for example, that training schools for military police must last at least eight weeks, shorter lengths of time being impractical for scheduling reasons. The solution would then have been narrowed to the second equipment configuration (B), category 3B and 4 soldiers, and a training time between eight and twelve weeks.

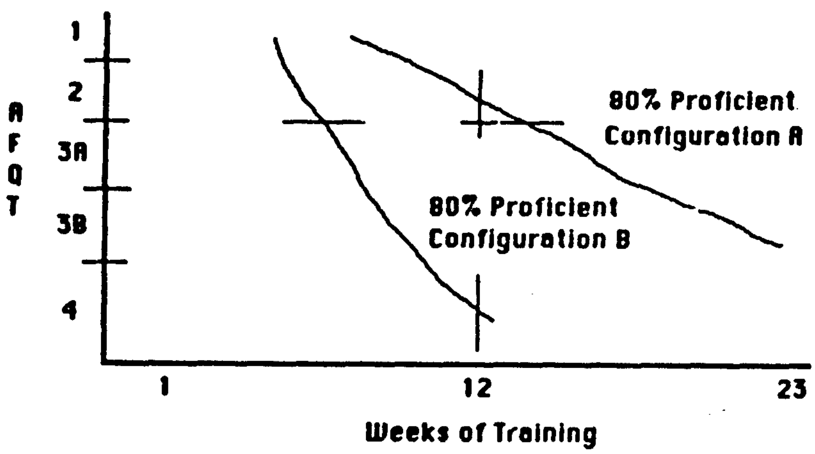

Figure 4. Figure 3 marked to indicate that category 1 and 2 soldiers are not available and that training times in excess of 12 weeks are too expensive.

Curves like the ones that appear in Figures 2, 3, and 4 can be generated in yet another way. Suppose two jobs are examined, one much simpler than the other. Figure 5 presents the situation. This time the two curves represent the same level of proficiency on two different jobs. Note that the curve for job A stretches out slowly to the right whereas the curve for job B drops much more sharply. Job A is aptitude sensitive. Any drop in aptitude level must be paid for by increased training time. Job B, on the other hand, is aptitude-insensitive. One can lower aptitude level without having greatly to increase training time. This difference has direct implications for personnel assignment. In any such situation one assigns high-aptitude soldiers to job A and lower aptitude soldiers to job B. The rule whereby one should proceed is clear. Starting with the high-aptitude end of the scale one assigns all soldiers to job A billets until they (the billets) are filled. Then one assigns the remaining soldiers to job B billets.

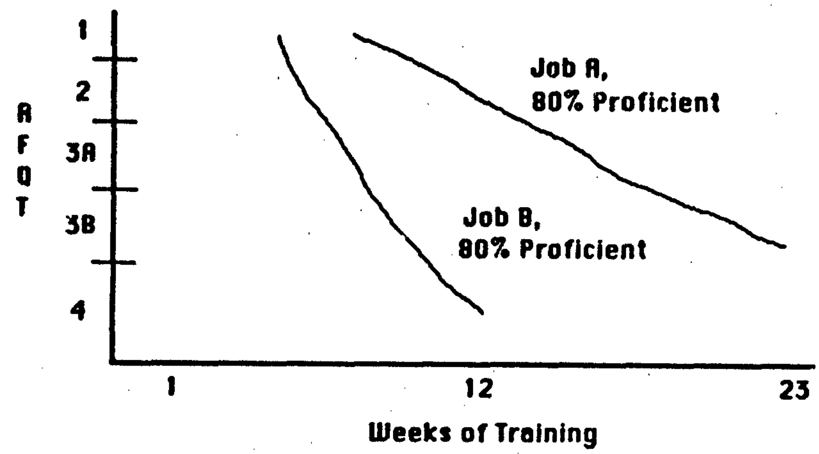

Figure 5. Two isoperformance curves representing the same level of proficiency (80%) on two different jobs.

The next section of the report describes the procedural set-up and analyses of "an illustrative experiment." The design of this experiment involved the variables of aptitude (gender), equipment (large versus small CRT screen), and training on a videogame task which simulated a remotely piloted vehicle. This study was successful as an isoperformance experiment because over 90% of the reliable variance is attributable to one of the three elements (training, subjects or equipment). Therefore, it can be employed both to illustrate how isoperformance methodology works and to examine "blocking-out" as a means of simplifying an isoperformance data set. The remaining sections will (1) discuss the application of the use of subject-matter experts and their role in isoperformance methodology in order to solve specific human factors engineering problems; (2) review specific applications to which isoperformance methodology can be put; and (3) describe broad suggestions for research and development work which should be accomplished.

AN ILLUSTRATIVE EXPERIMENT

Introduction

If user input is to be employed successfully in "performance reckoning" (or isoperformance), it must be kept simple and required on a limited basis only. If complicated or technical estimates are required, only a few people and perhaps none will be willing to make the effort. However, almost any experiment (ideal or not) involves many data points. The one that will be used to illustrate the isoperformance model involves only 24 subjects and a bare-bones design; yet even in this design systematic (nonerror) variance depends on 32 means. One way or another this number has to be reduced. An approximation has to be made. The moment, however, that one invokes approximation the question immediately arises as to how good that approximation is. One necessarily loses something when one approximates. The question is, how much?

The purpose of this illustrative experiment is twofold: (a) to describe one way ("blocking-out") that a set of experimental results can be approximated, and (b) then show how the "adequacy" of that approximation can be evaluated. The idea of adequacy will be developed formally later in this section but its general intent is to provide a quantitative index of how good an approximation is.

Task, Subjects, and Method

Task. The task used in this experiment is Air Combat Maneuvering (ACM) from the unmodified, commercially available Atari video game series. The task was designed to simulate a remotely controlled attack drone or RPV. The RPV task was implemented by an Atari Video Computer System (AVCS) on a Sears Model 564.5001 television with a 20-cm horizontal screen and a Sony Model KV-1917 television with a 45-cm horizontal screen. The subjects were seated approximately 0.6 m away resulting in displays of 19 and 43 degrees retinal size, respectively. The displays were generated in black and white on the TV screens after the Combat cartridge CX 2601, Task #24, was put into the AVCS and difficulty level B was set on the experimenter-controlled console. The task for the subject was to align a black, approximately triangular (1.3-cm base by 1.6-cm height) attack vehicle with a same-sized white target or drone jet moving at 5.5 cm/sec so that a fired missile would intercept. Experimenters initiated the task by pressing a reset button on the AVCS console. The subject controlled a joystick activated by the preferred hand on a control box with a "fire button" in the upper left-hand corner controlled by the nonpreferred hand.

Moving the joystick fore and aft increased or decreased the speed of the attack jet by 20%. Movements of the stick right and left turned the attack jet clockwise or counterclockwise at .67 rad/sec or approximately a rate sufficient to complete a 360-degree turn in four seconds. Combined lateral and vertical movements resulted in turns with changes in speed dictated by the joystick's vertical position.

Pressing the fire button launched a ballistic missile with an 11 cm/sec speed in a straight line with respect to the jet's body axis at the time of launch. If the missile intercepted the target jet, a hit was scored and the flight directions of both the target and attack jets were automatically rotated to new initial positions, 45 degrees clockwise and counterclockwise respectively. Further description of the AVCS and ACM task can be found in Atari (1977), and in Jones, Kennedy, and Bittner (1981). The military relevance of this task is evidenced by the fact that performance is highly correlated with performance on a full-scale simulation of the Navy's carrier landing task where corrected-for-attenuation correlations reveal more than 85% shared variance (Lintern & Kennedy, 1984).

The task is scored by recording the total number of hits at the end of each trial. Each trial is 2 minutes and 17 seconds long after which the game ends and is reset by the experimenter.

The equipment feature chosen was field of view measured by display screen size. Other variables could have been chosen (e.g., expert vs. novice settings). However, field-of-view size is a salient area in many current complex systems. For example, Westra and Lintern (1985), in simulated vertical takeoff and landing studies, obtained results indicating superior performance in helicopter hover landings with wide, as opposed to narrow, field-of-view.

The relevant aptitude measure is the gender of the subject. Gender, of course, is not itself a measure of aptitude. It happens, however, that men perform substantially better on almost all videogames than women (Jones, 1984). In this case, therefore, gender can be used to index aptitude in the same way that selection tests would.

Subjects. A total of 25 subjects were recruited for this study from the University of Central Florida. One subject attrited from the study, yielding a final *N* of 24. There were 15 female and 9 male subjects. All subjects signed a detailed informed consent form which explained the voluntary nature of participation, the types of tasks to be performed, as well as the compensation. Subjects were paid $5.00 a session for eight sessions.

Design and Procedure. The experimental design represents a mixed one in that aptitude and equipment (each at two levels) are group (between-subject) factors and sessions is a within-subject factor and crosses both aptitude and equipment. Each subject received five trials per day for eight days (Sessions) with no warm-up trials. On the initial day of testing all subjects were briefed about the procedure, tasks to be completed and a schedule for testing was arranged. Typically this consisted of coming in each weekday at the same time until finished.

Prior to the RPV task subjects were given a 2-3 minute briefing in which the task was described for the two conditions (big screen/small screen). Additionally, strategies were offered in an attempt to offset the large individual differences that were expected from prior experience with this and other video games. The strategies included; 1) "watch the drone as it flies off the screen and notice that it appears in exactly the same position on the other side of the screen." 2) "understand that your perspective is always as if you were flying the drone, so the vehicle will respond differently depending on your angle of attack" (subject was then shown that when the drone is coming down the screen from top to bottom and the joystick is moved to the right, the drone will turn to the left).

Results and Isoperformance Analyses

Results. Since the original experiment called for groups of equal size but, for reasons unrelated to the experiment (mainly the availability of subjects), the experiment was carried out with groups of unequal size, an unweighted-means analysis of variance is appropriate (Winer, 1971, p. 599). The alternative is a least-squares solution and is appropriate only if the groups represent strata within a specified population. This condition would hold for gender but not for the equipment variation, because using a large or small screen is an experimental condition and has no general application outside the present experiment. The allocation of subjects to groups was as follows:

Males, big screen	5
Males, small screen	4
Females, big screen	8
Females, small screen	7
Total	24

The unit of analysis was the average number of hits over the five trials within each of the eight sessions. Thus, 192 data points (8 X 24 subjects) were entered into the analyses.

Figure 6 presents the results for Aptitude (sex). The males do better than the females and by an amount that increases slightly with practice. Among the men the variance falls slightly late in practice; this is probably due to a ceiling effect.

Figure 7 presents the results for Equipment (big screen versus small screen). Clearly, the equipment variation in this case has no effect.

Figure 8 presents the results for Aptitude and Equipment, that is, for all four subject groups. Although, as will be seen, no statistical significance attaches to the result, there is a tendency for big vs. small screen to make more of a difference for females than for males. In fact, the males using the small screen did ever so slightly better than the males using the big screen.

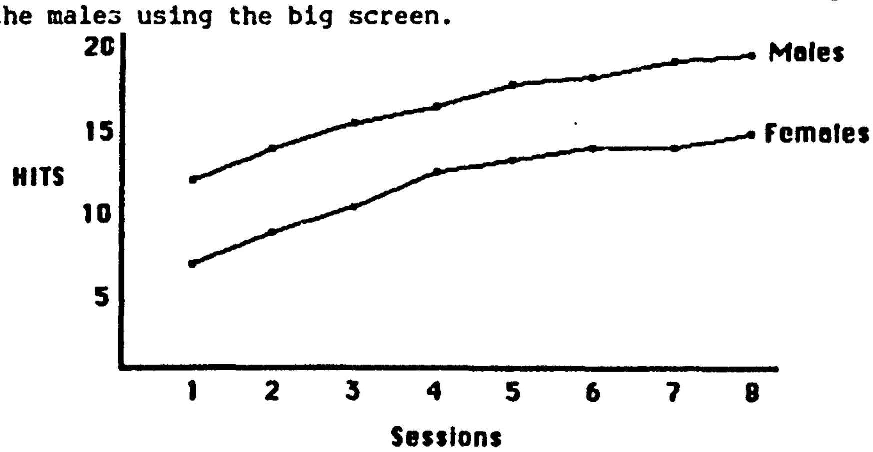

Figure 6. Average number of target hits over trials as a function of aptitude (gender) and session of practice: Unweighted means.

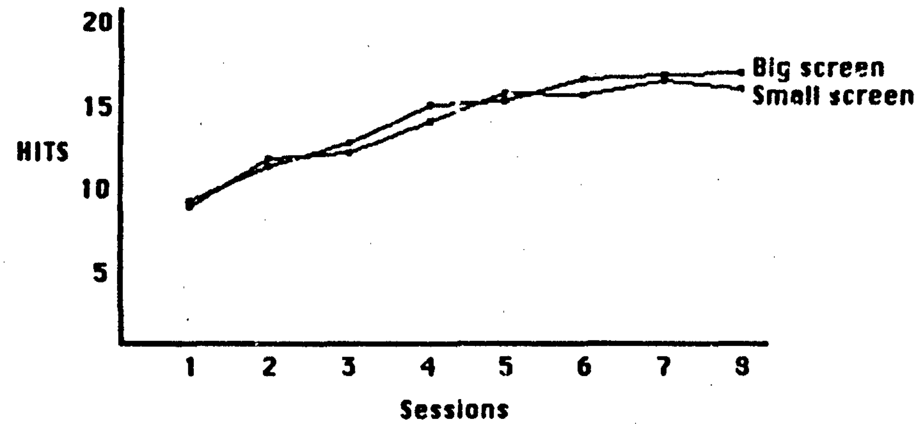

Figure 7. Average number of target hits over trials as a function of equipment and session of practice: Unweighted means.

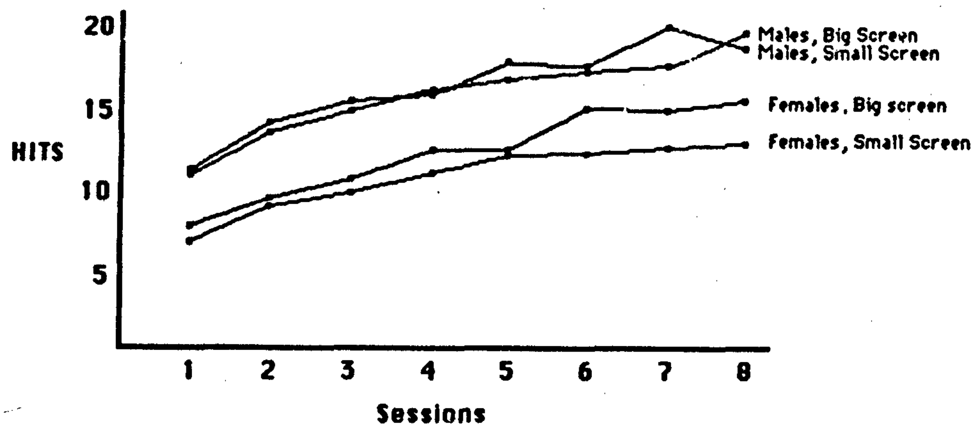

Figure 8. Avera : number of target hits over trials as a function of aptitude (gender), equipment, and session of practice: Unweighted means.

Table 1 presents the (unblocked) unweighted-means analysis of variance. The only significant effects are Gender ($F(1,20) = 22.8$, $p < .001$) and Sessions ($F(7,140) = 98.8$, $p < .001$). The appropriate error term for the first three components (A, E and AxE) is Subjects Within Groups, and for the next four components (T, AxT, ExT, and AxExT) is the Training-by-Subjects Within-Groups interaction. In an unweighted-means analysis the total variance does not in general equal the directly calculated total sum of squares. Therefore, the latter is not given (see Winer, 1971, pp. 599-602).

Table 1

Unweighted-Means Analysis of Variance

Source	SS	df	MS
Gender (A)	1,048.5	1	1,048.5
Equipment (E)	8.1	1	8.1
A X E	40.2	1	40.2
Sessions (T)	1,107.0	7	158.1
A X T	9.8	7	1.4
E X T	18.7	7	2.7
A X E X T	4.2	7	0.6
Subjects Within Groups	918.9	20	45.9
T X Subjects Within Groups	220.9	140	1.6
Total	-	191	-

Note: "A" = Aptitude; "T" = Training

<u>Blocking-Out the Experiment</u>. The purpose of blocking-out an experiment is to simplify a conventional, completely general design and analysis while accounting for as much systematic variance as possible. The procedure is to approximate a given data set with straight lines. One imposes on the data set a series of constraints which have this effect. In the present case we impose three constraints:

a. Practice is divided into two segments, early and late, each with four sessions;

b. All relations within segments must be linear;

c. No interactions are admitted within segments except Aptitude X Equipment.

In effect, this third constraint means that not only is practice segmented into linear components but so are its interactions with Aptitude and Equipment. Note, however, that Training X Aptitude and Training X Equipment interactions are reduced to zero only <u>within</u> segments. These interactions may still take nonzero values <u>between</u> segments. Hence, the blocked-out analysis will still include Aptitude X Training, Equipment X Training, and Aptitude X Equipment X Training interactions, albeit reduced by the removal of their within-segment components.

A blocked-out experiment is called an isoperformance model and the description of what follows should be compared with the theoretical predictions described in Figures 1-3 presented earlier in the paper. The isoperformance model consists exclusively of straight lines and not very many of them. In this case it consists of eight lines: performance as a function of Aptitude under either Equipment variation early in practice, and performance as a function of Aptitude under either Equipment variation late in practice.

An isoperformance model might not, of course, capture all or even the bulk of the systematic (nonerror) variance in the behavior of a military performance system. It is hypothesized, however, that it does. The total variance in performance can be divided into three mutually exclusive and collectively exhaustive parts:

- Systematic (nonerror) variance accounted for by the isoperformance model;
- Systematic variance not accounted for by the isoperformance model; and
- Error variance.

The "adequacy" of an isoperformance model is the proportion of the systematic variance in performance that it accounts for. To be acceptable, "adequacy" must be equal to or greater than 0.90. In Table 1 all components are systematic except the two error terms. The question now is, how much of this systematic variance (or sum of squares) can be captured by the isoperformance model described above?

<u>Fitting the Straight Lines</u>. The requirement of no interaction with training within segments means that all performance functions within segments must be parallel. In this case, for example, the four subject groups (males and females by big and small screens) must all follow parallel courses over Sessions 1-4. Similarly, they must follow parallel courses in the second segment over Sessions 5-8. However, the slopes of the two sets of parallel lines do not have to be the same, nor must the differences among the four groups be the same in the two segments.

Consider now either one of the segments, say, the first. We wish to fit the following linear regression model,

$$Y_{ij} = \alpha_i + \beta(X_j - \bar{X}) + \epsilon_{ij}$$

where Y_{ij} is performance of the <u>ith</u> group (i = 1,..., 4) on the <u>jth</u> session (j = 1,..., 4), α_i is the intercept for the <u>ith</u> group, (is the slope of all four lines, and X_j is session number). The final term, ϵ_{ij}, is the error term for the <u>ith</u> group in the <u>jth</u> session and is assumed to be normally distributed with mean zero.

In the case of the present experiment (first segment), beta = 1.67, and α_1 = 14.3, α_2 = 14.7, α_3 = 10.2, and α_4 = 9.3, where α_1 refers to males using the big screen, α_2 to males using the small screen, α_3 to females using the big screen, and α_4 to females using the small screen. For the second segment, beta = 0.45, and α_1 = 18.4, α_2 = 19.1, α_3 = 14.7, and α_4 = 12.9.

Note the approximately fourfold decrease in slope from the first to the second segment, as well as the increases in intercepts (average values).

Figure 9 presents the blocked-out results graphically. The next step is to carry out an analysis of variance on these "data."

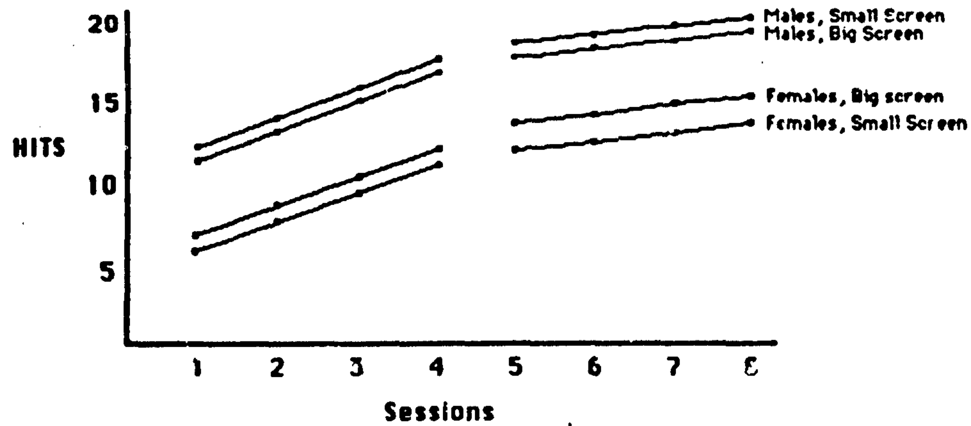

Figure 9. Blocked-out means of the average number of target hits over trials as a function of aptitude (gender), equipment, and session of practice. Each line represents a linear regression model of the data in Figure 8.

The Blocked-out Analysis of Variance. The blocked-out analysis of variance was carried out using the 32 blocked-out means in Figure 9 as data points for calculating systematic (nonerror) sources of variance rather than the corresponding unblocked data points in Figure 8. Table 2 presents the blocked-out analysis. Note, first of all, that the first three components (A, E, and AxE) are exactly the same as in the unblocked analysis. That is because blocking-out leaves the means of the four groups exactly as they were. The sum of squares for sessions is a trifle smaller than in the unblocked analysis because mean performance is not perfectly accounted for by two straight lines (early and late). One might think that the next three components should all equal zero, because all interactions within segments are ignored in the isoperformance model; and within segments these interaction components do, in fact, vanish. However, for practice as a whole (both segments), they do not vanish because the differences among subject groups are not necessarily the same in the two segments. The males, for example, have a 4.78 point edge on the females in the first segment and a slightly larger edge, 4.92, in the second segment. Thus, the two curves, though parallel within segments, are not parallel throughout practice. The same is true for equipment.

Table 2

Blocked, Unweighted-Means Analysis of Variance

Source	SS	df	MS
Gender (A)	1,048.5	1	1,048.5
Equipment (E)	8.1	1	8.1
A X E	40.2	1	40.2
Sessions (T)	1,101.0	7	157.3
A X T	0.3	7	0.1
E X T	1.0	7	0.1
A X E X T	3.3	7	0.5
Total (Systematic)	2,20 .5	31	-

Note: "A" = Aptitude and "T" = Training

The blocking-out process concerns systematic (nonerror) variance only. It depends only on the variance among the means of the four groups over the eight practice sessions; sums of squares and mean squares for Subjects Within Groups and Training X Subjects Within Groups are not involved in the blocking-out analysis and are, therefore, omitted in Table 2.

Adequacy of the Model. In blocking-out an experiment some variance is lost in the form of deviations of the empirical data from the straight lines used to block-out the experiment. How large do these deviations loom? How much variance (or sum of squares) do they represent? In the unblocked data systematic sums of squares totalled 2,236.5 and in the blocked-out analysis systematic components totalled 2,202.5. Therefore, the adequacy of the isoperformance model is

$$\frac{2{,}202.5}{2{,}236.5} = 98.5\%$$

In this case, it is possible to greatly simplify the data set (i.e., from 32 means into 2 slopes and 8 intercepts) at a trifling cost in lost variance.

Isoperformance Curves. In order to obtain isoperformance curves one must first decide on a level of performance that constitutes "proficiency." For purposes of illustration, a score of 13 was used as the cut-off point for proficiency. That is, the specified operational requirement for a suitable RPV operator is to obtain 13 hits in a 2.28-minute time frame (the period of performance of one trial for an Atari AVCS game #24 program). In the unblocked data (Figure 8) males achieved this level using the big screen after 1.6 sessions and females after 3.8 sessions. Using the small screen, males reached a mean score of 13 after 1.4 and females after 7.0 sessions. The isoperformance curves, therefore, for the unblocked data take the form shown in Figure 10. Any combination of Aptitude, Equipment, and Practice shown on either of these curves will produce a group of people with mean performance equalling 13. If one chooses males one can achieve this level quicker than if one chooses females. The A X E X T interaction is larger for females than for males, though not significantly so. Using the small screen might be cheaper or have other advantages (for example, lower weight or less volume) but these advantages might well be outweighed by the substantially greater amount of time (and money) needed to train women to a level of proficiency. Men, of course, could also be used but they, being the higher aptitude group for this sort of job, might be needed elsewhere.

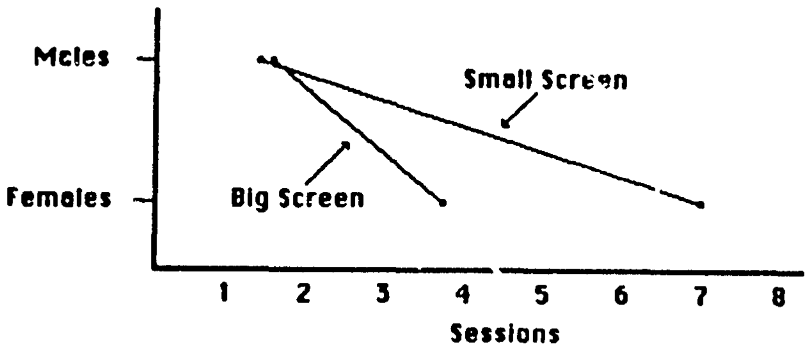

Figure 10. Isoperformance curves for the unblocked means: Number of sessions required to reach a proficiency of 13 target hits per trial.

Figure 11 presents the isoperformance curves for the blocked-out data, using the same cut-off point. Plainly, blocking-out does not disturb the isoperformance curves appreciably.

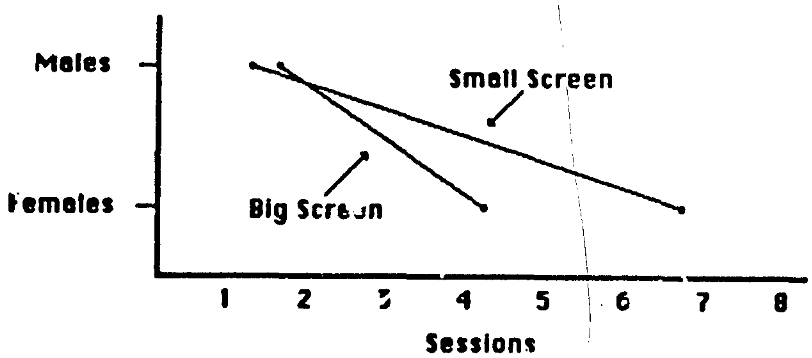

Figure 11. Isoperformance curves for the blocked-out means: Number of sessions required to reach a proficiency of 13 target hits per trial.

THE ROLE OF SUBJECT-MATTER EXPERTS

The blocking-out process described in the preceding section is designed to simplify a design. But what, one may ask, is the purpose of such a simplification? What good does it do to reduce the number of parameters needed to circumscribe the systematic variance in a data set? If one has empirical results, as in the illustrative experiment, it does no good; but in many of the most important situations from an applied point of view one does not have empirical results or at least not full results and, as a consequence, one must extrapolate from earlier and similar situations to the one at hand. In these cases simplification is helpful. The fewer the number of estimates that have to be made the easier it is to make them and to make them with reasonable accuracy.

In the design of a new weapon system, for example, one cannot empirically determine the human factors requirements of the new system because that system does not yet exist. Even if it did, an empirical determination is likely to be out of the question for practical reasons. It could be inadmissibly expensive to train various categories of personnel to proficiency simply to find out how long it takes. The present state of systems research and human factors science, however, does not allow a strictly deductive application of general principles to these problems. Human-factors science simply is not that advanced. A great deal, of course, is known but not enough to provide clear-cut answers on a strictly deductive basis (that is, not involving human judgment) to such questions as how aptitude and training time will trade off in a new weapon system. To a certain extent a human being must extrapolate from known results regarding existing systems to the new system. That being the case, any attempt completely to simulate the aptitude-by-training-by-equipment trade-off is bound to be arbitrary. One can do it, of course, but any particular simulation has no claim on our attention that an infinity of different simulations would not also have.

An alternative is a decision to simulate the warrantable science available and to use human judgment, or more explicitly subject-matter experts (SMEs), when existing evidence must be extrapolated to a new system. In this report isoperformance methodology is similar to HARDMAN and similar approaches. The next order of business is which particular judgments the subject-matter experts should make. Three general points are clear at the outset. First, the judgments that the SMEs make should not be technical. Only the simplest and most familiar ideas should be used. Second, the amount of user estimation should be held to a minimum. Third, the judgments made by the SMEs must be sufficient, together with the warrantable science built into the methodology to generate isoperformance curves.

Before discussing the role of blocking-out in shaping the judgments to be made by SMEs, it will be necessary to digress briefly regarding related work. Concurrently with the present contract, an effort has been ongoing under Air Force auspices to develop an interactive computer program to implement isoperformance methodology (Jones et al., 1987). The general idea is to write a program that will allow a relatively unsophisticated person to use isoperformance methodology effectively. User input will be as minimal and as simple as possible. Libraries of relevant training and aptitude information will be made accessible to the user, and checks based on warrantable science will be built into the program. The output of the program will be isoperformance curves. In the remainder of this report the effort to write such an interactive computer program is assumed and much of the discussion will revolve around it.

One more matter needs to be addressed before continuing with the discussion of user input. It concerns the validity of the isoperformance approach and how it can be determined. As already noted, isoperformance is intended to be used primarily in situations where human-factors requirements must be projected for a system which does not at the time exist. How is it known that these projections are correct? Granted that projections must be taken for a new system, hence, how much confidence can we have that the isoperformance program allows us to make reasonably accurate projections? The answer is that the isoperformance program must be tested against empirical results in situations which do exist. If it is accurate there, then a basis is formed for expecting it to be accurate in situations where no test is possible. A methodology for validating the isoperformance approach has been worked out and, hopefully, will be implemented in the near future. The details of this methodology need not be of concern here. It is, however, worth pointing out that the proposed methodology allows the validation of specific parts of the isoperformance program (for example, the training and aptitude libraries) as well as the program as a whole. A means, therefore, has been developed for knowing which parts of the program work well and which do not. This, in turn, allows not only a validation of the program but to pinpoint where it is not working and to improve it.

What is the role of user input in the isoperformance program and how should that input should be shaped? The illustrative experiment described earlier clearly suggests that blocking out is one way of simplifying an experiment so as to reduce the extent and complexity of the estimates a user has to make. In that experiment the systematic sources of variance depended on 32 empirically determined values, specifically the means of the four subject groups over the eight sessions of practice. In the absence, therefore, of blocking-out, it would be necessary for a user to make 32 estimates. Blocking-out reduces this number to 10 estimates, namely the slope and four intercepts in the two segments. This is a substantial simplification and, as has been seen, one that can be achieved with little loss of variance (information). But is it a sufficient simplification to allow user input to be made in these (blocked-out) terms? Perhaps, but there are several reasons for concern. First, 10 estimates are probably still too many. Second, the idea of a mean may be widely understood but that of a slope is not. Third, both means and slopes depend on a particular performance measure and its units.

This last point is the most troublesome. In order to estimate slope, for example, one has to specify how many units on the performance measure a given group of subjects will improve each session. This is by no means a simple task. To begin with, few users are likely to be familiar with the particular tests or exercises used to evaluate performance at the end of training. If one insists on the user's knowing these things, then the pool of potential users will be limited to a handful of personnel experts. The units on the performance measure are another problem. How many people know what they are and can think intelligently in terms of them? One could, of course, resort to standard scores but then the user would have to know what a standard deviation is. Again potential users would be limited to a relatively few technically knowledgeable people.

In the isoperformance computer program currently under development, user input is made in terms of personnel categories, percentages, proficiency, and amounts of time, all of them widely and easily understood terms. The term "personnel categories" simply means a group of people: men, women, Mental Category 2 soldiers, average or normal high spatial frequency visual contrast sensitivity, the top 10% on mechanical ability, etc. Percentages, along with categories, allow one to avoid not only means but also the performance measure and its units. Instead of estimating means on a quantitative measure, the user estimates percentages of soldiers in a given category who are proficient. Proficiency is another way of avoiding estimates in terms of quantitative performance. Instead of estimating numerical values (means, increments with practice, etc.) one estimates percentages of soldiers who meet a minimum standard. That standard is, of course, implicitly defined on a performance measure. Nevertheless, one can specify percent proficient without having to specify numerical results and, in fact, all of the military services do just that.

Blocking-out, however, is a useful procedure and one that certainly has a future in isoperformance methodology. It may be used to help shape user input and could easily find a place in a subprogram on retention and transfer or, perhaps, in a tutorial subprogram.

SPECIFIC APPLICATIONS

Isoperformance methodology has broad application in systems research for government and private industry. Five major areas are (a) as a management decision aid for human factors engineering design; (b) as an adjunct to aid in organizing manpower, personnel, and training (MPT) applications, particularly where "what if" questions need to be answered and where an audit trail of the solution adopted is useful; (c) as a formal system for conducting trade-offs where cost analyses are conducted for existing systems; (d) as a means of implementing the recent DoD policy mandating the use of Nondevelopmental Items (NDI); and (e) as a way for industry to meet the functional specifications and requirements in an RFP. A brief example is provided for each of these areas to demonstrate the utility of isoperformance methodology.

Isoperformance kinds of estimates are already required in the form of MANPRINT analyses. The data available from the MANPRINT requirements for systems will work well as data for explicit trade-offs in isoperformance analyses. These types of trade-offs among aptitude, training, and equipment are the type DoD has requested for the last few years.

Within the MPT arena, isoperformance methodology permits trade-offs for each component and provides immediate feedback for forcasting efficiency and selection/placement. The current IsoDemo program developed for the Air Force (Jones & Jones, & Essex Corporation, 1987) provides a constrained example of using isoperformance methodology for selection and placement. The program provides an example for a jet mechanic's position. When the training time and percent proficient within that training time are specified, the program generates isoperformance curves. From these curves the aptitude category necessary (ASVAB, AFQT) to fill that mechanic position can easily be seen. If other equipment or flexibility in the training schedule is available, these estimates may change and feedback is immediate. Importantly, a record of all the options can be obtained within a short period.

At the end of the program the MPT specialist or manager can tell what the lowest aptitude category is within the training time and equipment constraints available. Conversely, he/she can also find the minimum training time necessary if the very best people were available as one would hope in private industry selection.

The third major area of isoperformance has the broadest application. This area is using isoperformance methodology for existing systems. Isoperformance can be used to evaluate and suggest improvements in any system where there is a man/machine interaction or the various costs of the different parts can be compared. This is especially useful with emerging technologies. Tice (1986) for example has pointed out that the Army's Stinger weapon system was unable to be operated properly when fielded because allowance had not been made for the differences in visual capability of the operators. Additionally, on the micro-level of a single system, a recent study concerning the F-15 Eagle fighter plane (Dedrick, 1986) it was noted that, "A critical assumption is made that pilot proficiency is keeping pace with the rush of rapid technological improvements" (p. 37). In conclusion the report stated that: "The current trend in aircraft capability analysis is to overemphasize hardware. Emphasis must be placed on the complete weapons system of the man and machine when evaluating our warfighting capabilities" (p. 37). The situation with the F-15 is an ideal example for application of isoperformance methodology. It would force the user to see the high level of aptitude (in this case knowledge of equipment, flying skills, and flying time) required for the aircraft as well as the long training times. From this analysis specific areas could be targeted for intervention, such as automating certain functions of the aircraft to lower the aptitude requirements and training time. Additionally, and again, there is an audit trail of the various decisions which were made.

On a macro level the isoperformance approach is well suited for application of the recent DoD policy mandating the use of Non-Developmental Items (NDI) in the acquisition process. This NDI procurement plan is a direct result of the President's Council on Defense Acquisition, the Packard Commission. Governmental agencies are required to evaluate the ability of an "off-the-shelf" item for satisfying their functional needs. An NDI may be entirely off-the-shelf needing no development or the item may require a dedicated R&D effort by the contractor to modify the item for current governmental needs. A major principle in NDI acquisition is that less than full compliance with a programs performance objectives is insufficient reason not to use NDI. In other words, if an NDI does not meet all specifications and requirements set forth in the Request for Proposal (RFP), it is not disqualified; cost/benefit trade-offs can be made. Here lies the isoperformance strong point. In NDI acquisitions isoperformance techniques can be used by the Acquisition Review Board to check a program manager's (PM) choice of NDI or R&D program. The NDI will have data available and estimates may be gathered for the R&D program much as in system design. Similarly, the PM can assess the current manpower and training situation to see if an item fits the user's needs with realistic demands on the labor pool and training school.

Finally, industry may use isoperformance methodology to meet the functional specifications and requirements in an RFP. Suppose the government calls for an NDI acquisition for updating or replacing an in-place piece of equipment. A company may propose to modify the system by upgrading it to make it "state-of-the-art," or it can trade-off the complexity through longer training time or selection of higher aptitude personnel. The company may propose to replace the equipment with a less complex system with no development cost associated. In this way the company cannot only lower the unit cost but could provide isoperformance verification for shorter training time and broader use of the labor pool. This would result in substantial lowering of total system costs in training, personnel and probably integrated logistic support (ILS) and reliability and maintainability data (RAM) costs. The benefits are obvious, the company may elect to pursue a technological advantage or an overall cost advantage. Both are defensible and may be suggested to a program manager for overall preference. If the system is a trainer or simulator, state-of-the-art may be required. If it is a vehicle an overall cost approach may be chosen. The Army, for example, adapted the Chevy Blazer to meet their light truck requirements.

As a computerized decision aid in design, the isoperformance program may be used to trade-off the aptitude, equipment, and training dimensions which are known or can be estimated for a prospective system. In this way overall utility as well as cost/benefit considerations may be assessed. For example, in a new weapons system, the projected manpower of the target service as well as the allowable minimum and maximum training times may be reasonably estimated. This will form a "window" within which the equipment (man/machine interface) must stay. Many questions about which elements to emphasize can be answered almost immediately by framing the question within the context of the isoperformance model.

ADDITIONAL RESEARCH

A logical extension to this Phase 1 effort would be (a) to develop and computerize as interactive programs three key components within the isoperformance package, and (b) to validate the technical venture as a whole and several main parts within it against appropriate empirical results. These two general objectives will be discussed in the order stated.

Program Development (IsoTutor, IsoApply, IsoEquip)

Figure 12 presents a flow chart for an overall isoperformance computer program package as currently conceived. When the main program comes up, the user has three options. The first, IsoDemo, is an orientation program. It explains and illustrates the main ideas of isoperformance methodology. This program is primarily didactic in nature and has already been written. It will, no doubt, go through one or two revisions but the principal features of the program are not likely to change. IsoCore is scheduled for completion under Air Force contract by the end of calendar 1988. IsoCore is the central working subprogram in the package. In it the user makes estimates for training times and percent proficient for different aptitude categories. These estimates have been deliberately couched in terms that are almost universally understood: categories, percentages, and amounts of time (not means, standard deviations, correlations or more complex statistical ideas). In addition, the number of estimates has been greatly reduced by an optional "expedite" procedure. In this procedure the user makes estimates for the top and bottom categories and the program "fills in" the estimates for the intervening categories under the assumption that performance is linearly related to the aptitude dimension specified in defining aptitude categories. The user has the option of making more detailed estimates or correcting the ones generated by the expedite procedure, but reasonable input curves can be generated with as few as five estimates. In making these estimates the user has access to a Training Library which contains relevant information about related jobs. How long does it take to train soldiers for jobs similar to the ones being considered? What aptitude levels are required for entry into a school that provides such training? How much prior experience is required? In making estimates it is assumed that the user, in effect, knows whatever is available to be known about similar kinds of training. Once made, the estimates are checked, among other ways, by means of the Validity Library. It is possible to derive an implicit correlation between the specified aptitude dimension and performance at the end of training from the user's input estimates. Such a correlation is a predictive validity, and a great deal is known about the predictive validities of various aptitude dimensions for most training programs. The Validity Library is a compendium of these predictive validities and it allows the user to check the predictive validity implicit in his or her input estimates against known values for similar programs. If the user's implicit validity is out of line with these values, a correction is in order. The Training and Validity Libraries are the principal data bases for the isoperformance program package. IsoCore eventuates in isoperformance curves that describe all combinations of aptitude level and training time that produce the same percentage of proficient soldiers. The user is able to specify any desired level (50%, 70%, 90%, or whatever). Thus, IsoCore eventuates in a family of isoperformance curves.

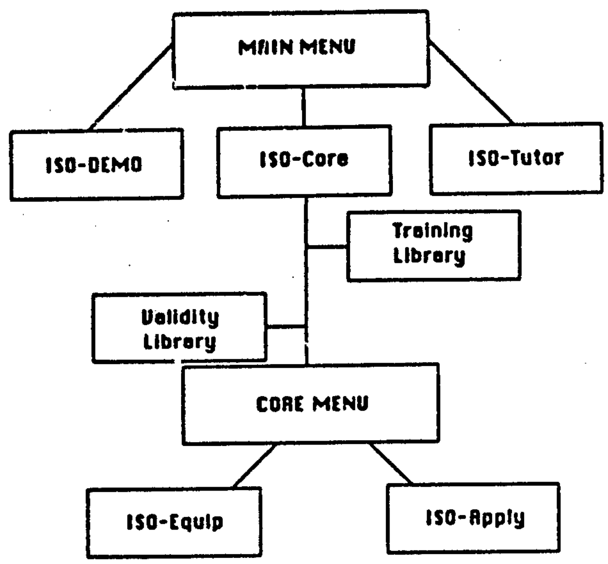

Figure 12. Current flow chart for the isoperformance program package.

In our judgment, future extension of this work in three crucial directions is advisable. The first is IsoTutor. If it is to be effective, the isoperformance program package must be accessible to as broad a range of potential users as possible. For this reason the ideas required for user input have been limited to categories, percentages, and amounts of time, as already noted. It must be anticipated, however, that some users may be mid-level military or civil servants with responsibilities for the acquisition of military systems and may not only not be familiar with systems research and human factors, but may have little or no background in psychology or in the study of skill acquisition. IsoTutor has been conceived with these potential users in mind. The subprogram will be built about a videogame that simulates a remotely piloted vehicle or some other militarily relevant task which is easily represented with a microcomputer. The user will be asked to imagine that controlling this drone is the task to be learned. The user will then practice the videogame. In this hands-on manner the user will be brought to understand a series of very basic truths about skill acquisition, for example, that learning curves are generally negatively accelerated, or that a given category of soldiers after a fixed amount of practice do not all perform at the same level but distribute themselves in a broadly "normal" way about a central value, or that an isoperformance curve shifts downward and to the left if the job is made easier by automating parts of it. The purpose of IsoTutor is to bring managers with insufficient background or potential subject-matter experts "up to speed" on the empirical content of the isoperformance package. It may also be helpful in articulating discussion intended to implement the MANPRINT mandate. Rational values (time and money) can be used for equipment, training, and selection costs, after which the program can be exercised. For some users it will function as a confidence builder in making estimates. The Training and Validity Libraries also serve this function.

When the user arrives at the Core Menu, he or she will have isoperformance curves for a given equipment configuration, but these curves will not yet have been "evaluated." All points on an isoperformance curve produce the same level of performance, but some of these points call for soldiers who are in great demand for other jobs, other points require exorbitantly expensive training times, still others are administratively infeasible because they do not conform to existing procedures or are incompatible with existing structures. IsoApply assists the user in narrowing down an isoperformance curve to a few points or ranges of points that can be recommended for adoption by the Army. The user needs to be aware, for example, that mixes of students, some lying above the isoperformance curve aptitudinally and some below it, can be recommended provided the numbers of students above and below the curve are balanced. One further needs to know specific mixes of students which meet this requirement. IsoApply assists the user in all these respects.

The possibility cannot be excluded, however, that a given equipment configuration may not allow any satisfactory solution. In such a case the possibility of equipment redesign can be considered. Perhaps the job can be partly automated so that it can be done by lower aptitude personnel or with less training. Obviously, such information is available and can be iterated and could provide important management information which could serve to improve arguments to legislative and budget control agencies regarding military systems. It is at this point that IsoEquip enters the picture. If a second equipment configuration is to be considered, one possibility is simply to specify it and repeat IsoCore. Here again, however, an "expedite" procedure can be developed. Suppose that the new configuration simplifies the job. If so, it can be equated, at least provisionally, with the original configuration and a lower cut-off point for determining what is satisfactory performance (proficiency). This shift of the cut-off point, however, can be accomplished in a single estimate and, once made, allows a complete set of isoperformance curves to be drawn. The user may then make adjustments in these curves if so desired. Therefore, these can be employed to soften impacts on the dwindling manpower pool (Merriman & Chatelier, 1981). Once a second set of isoperformance curves has been decided upon, the user may return to IsoApply for further evaluation.

Validity Study

A validity study, to be carried out in subsequent work, would be, broadly speaking, a deletion experiment. Major components of the program, for example, the Training or Validity libraries, would be taken out and the accuracy of the program with and without these components compared. For example, four groups of subject-matter experts could use the program. One group would have access to the entire program. A second group would have access to the Validity Library and other checks on user estimates but not to the Training Library. A third group would have access to the Training but not to the Validity Library. The fourth group would not have access to either library. If, as hypothesized, user estimates more closely approximate real-world results the more complete the program is, the fact would argue strongly for validity. If a program's components improve validity, then the program itself must be valid, at least to the extent of the improvements. Comparisons with altogether different approaches are more difficult to come by. If, however, any such approach turns out to be feasible, an experimental design comparing it to the isoperformance approach would not be difficult.

SUMMARY

Pressures of budgets and increasing technological sophistication imply that cost/benefit trade-offs need to be examined, and blue-ribbon panels advocate the kinds of trade-offs which Isoperformance Methodology is designed to make. More recently DoD elements have mandated various programs to accomplish such ends.

This report provided empirical support in the form of an experiment for the Isoperformance Methodology, and delineated the functions for a "smart" interactive computer program for human factors decision making in systems research. The report also addressed key technical issues and how they would be handled in order to prosecute such a program.

The continuing development of such work would provide a computerized decision aid to be employed as a managerial tool for systems design, evaluation of in place systems, and MPT planning including defensible selection and placement practices.

Additionally, it is believed that Program Managers and Acquisition Review Boards in DoD could use Isoperformance Methodologies in assessing the benefits of pursuing Nondevelopmental or standard R&D procurement strategies and subsequent justification of those choices.

REFERENCES

Alluisi, E. A. (1978). Performance measurement technology: Issues and answers. Norfolk, VA: Performance Assessment Laboratory.

Atari, Inc. (1977). Combat game program instructions. Sunnyvale, CA: Atari Inc., Consumer Division (C011402-01).

Ault Committee Report. (1969). Report of the air-to-air missile system capability review. Washington, DC: COMNAVAIRSYSCOM.

Berg, C., Hertzog, C., & Hunt, E. B. (1982). Age differences in the speed of mental rotation. Developmental Psychology, 18, 95-107.

Bittner, A. C., Jr., & Moroney, W. F. (1984). Estimation of comprehensive anthropometric data from limited information. Proceedings of the 28th Annual Meeting of the Human Factors Society (pp. 176-178). San Antonio, TX: Human Factors Society.

Bittner, A. C., Jr., & Moroney, W. F. (1985). The accommodated proportion of a potential user population: Compilation and comparisons of methods for estimation. Proceedings of the 29th Annual Meeting of the Human Factors Society (pp. 376-380). Baltimore, MD: Human Factors Society.

Boff, K. R. (Ed.). (1984). Handbook of perception and human performance. Wright-Patterson Air Force Base, OH: Air Force Armstrong Aerospace Medical Research Laboratory.

Campbell, A. E. (1970). Red Baron I summary: Visual detection in air-to-air combat (Technical Note 4011-2: Doc. No. 40-1327). China Lake, CA: Naval Weapons Center.

Chapanis, A., Garner, W. R., Morgan, C. T., & Sanford, F. H. (1947). Lectures on men and machines: An introduction to human engineering (Report No. 166-I-19). Baltimore, MD: The Johns Hopkins University, Department of Psychology.

Daneman, M. (1983). The measurement of reading comprehension: How not to trade construct validity for predictive power. Intelligence, 6, 331-345.

Dedrick, C. R. (1986). Effect of current rated personnel management policies on the combat capability of the F-15 aircraft (Rep. No. 86-675). Maxwell AFB, AL: Air Command and Staff College.

Department of Defense. (1981). Military Standard 1472C. Human engineering design criteria for military systems, equipment, and facilities (MIL-STD-1472C). Washington, DC.

Fitts, P. M. (Ed.). (1951). Human engineering for an effective air investigation and traffic control system. Washington, DC: National Research Council.

Green, B. F., & Hall, J. A. (1984). Quantitative methods for literature reviews. Annual Review of Psychology, 35, 37-53.

Hagman, J. D., & Rose, A. M. (1983). Retention of military tasks: A review. Human Factors, 25, 199-213.

Harbeson, M. M., Bittner, A. C., Jr., Kennedy, R. S., Carter, R. C., & Krause, M. M. (1983). Performance Evaluation Tests for Environmental Research (PETER): Bibliography. Perceptual and Motor Skills, 57, 283-293.

Harris, R. M., Iavecchia, H. P., Ross, L. V., & Shaffer, S. C. (1987). Microcomputer human operator simulator (HOS- IV). Proceedings of the 31st Annual Meeting of the Human Factor Society (pp. 1179-1183). Santa Monica, CA: Human Factors Society.

Hays, W. (1977). Statistics for psychologists. New York, NY: Holt, Rinehart, & Winston.

Hunt, E. (1984). Intelligence and mental competence. Naval Research Reviews, 37-42.

Hunt E., Davidson, J. E., & Larsman, M. (1981). Individual differences in long-term memory access. Memory and Cognition, 9(6), 599-608.

Jones & Jones, & Essex Corporation. (1987). A brief introduction to Isoperformance methodology [Computer program]. Orlando, FL: Essex Corporation. (Armstrong Aerospace Medical Research Lab. Contract No. F33615-86-C-0553)

Jones, M. B. (1984). Video games as psychological tests. Simulation Games, 15, 131-157.

Jones, M. B., Kennedy, R. S., Baltzley, D. R., & Westra, D. P. (in preparation). Omega square analyses and reevaluation of a meta-analysis for equipment features and individual differences in an aircraft simulator. Orlando, FL: Essex Corporation.

Jones, M. B., Kennedy, R. S., & Bittner, A. C., Jr. (1981). A video game for performance testing. American Journal of Psychology, 94, 143-152.

Jones, M. B., Kennedy, R. S., Kuntz, L. A., & Baltzley, D. R. (1987). Isoperformance: Trading off selection, training, and equipment variations to maintain the same level of systems performance. Proceedings of the 31st Annual Meeting of the Human Factors Society (p. 634-631). Santa Monica, CA: Human Factors Society.

Jones, M. B., Kennedy, R. S., Turnage, J. J., Kuntz, L. A., & Jones, S. A. (1986). Meta-anlysis of human factors engineering studies comparing individual differences, practice effects, and equipment design variations (Final Rep. No. F33615-85-C-0539). Wright-Patterson AFB, OH: Department of the Air Force, Air Force Systems Command, Aeronautical Systems Divisions. (NTIS No. AD A168 790)

Jordan, N. (1963). Allocation of functions between man and machines in automated systems. Journal of Applied Psychology, 47(3), 161-165.

Kearns, J. H. (1982). A systems approach for crew station design and evaluation (AFWAL-TR-81-3175). Wright-Patterson AFB, OH: Flight Dynamics Laboratory. (NTIS No. AD A115 663)

Kennedy, R. S., Berbaum, K. S., Collyer, S. C., May, J. G., & Dunlap, W. P. (1983). Visual simulation requirements for aircraft aspect recognition at real-world distances (NAVTRAEQUIPCEN 81-C-0105-5). Orlando, FL: Naval Training Equipment Center.

Kennedy, R. S., Bittner, A. C., Jr., Harbeson, M. M., & Jones, M. B. (1982). Television computer games: A "new look" in performance testing. Aviation, Space and Environmental Medicine, 53, 49-53.

Kyllonen, P. C., Woltz, D. J., & Lohman, D. F. (1981). Models of strategy and strategy shifting in spatial visualization performance (Technical Rep. No. 17, Aptitude Research Project), Stanford, CA: School of Education, Stanford University

Lane, N. E. (1986). Skill acquisition curves: An analysis of applicability to the time course and sequencing of military training (Final Rep., Contract No. 5540, Institute for Defense Analysis). Orlando, FL: Essex Corporation.

Laughery, R., Drews, C., Archer, R., & Kramme, K. (1986). MicroSAINT simulation analyzing operator workload in a future attack helicopter. In NAECON 86 Meeting (pp. 896-903). Dayton, OH: NAECON.

Lintern, G., & Kennedy, R. S. (1984). A video game as a covariate for carrier landing research. Perceptual and Motor Skills, 58, 167-172.

Lintern, G., Nelson, B. E., Sheppard, D. J., Westra, D. P., & Kennedy, R. S. (1981). Visual Technology Research Simulator (VTRS) human performance research. Phase III (NAVTRAEQUIPCEN 78-C-0060-11). Westlake Village, CA: Canyon Research Group, Inc.

Maier, M. H., & Grafton, F. C. (1980). Renorming ASVAB 6/7 at Armed Forces examining and entrance stations (Technical Memorandum 80-1). Washington, DC: Office of the Assistant Secretary of Defense.

Malone, T. B., Shenk, S. W., & Moroney, W. F. (1976). Human factors test and evaluation manual (HFTEMan, Vols. 1-3 (Tech. Rep. No. N00123-76-C-1364). Point Mugu, CA: Pacific Missile Test Center.

Mannle, T. E., Guptill, R. V., & Risser, D. T. (1985). HARDMAN comparability analysis methodology guide (Vols. 1-5). Wilmington, MA: Dynamics Research Corporation.

Marcus, A., & Kaplan, J. D. (1984). Reverse engineering of the MI fault detection and isolation subsystem: Human factors, manpower, personnel, and training in the weapons system acquisition process (ARI Research Note 84-101). Alexandria, VA: U.S. Army Research Institute. (NTIS No. AD A142 850)

Merriman, S. C., & Chatelier, P. R. (1981). Impact of the Department of Defense science and technology program on training and personnel systems technology. Washington, DC: Office of Environmental and Life Sciences. (NTIS No. AD A103 382)

Morgan, L. T., Cook, J. S., III, Chapanis, A., & Lund, M. W. (Eds.). (1963). Human engineering guide to equipment design. New York: McGraw-Hill.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.). Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Palmer, J., McLeod, C., Hunt, E., & Davidson, J. (1983). Information processing correlates of reading. Unpublished paper. Seattle, WA: University of Washington Psychology Department.

Perkins, J. C., Binel, D. C. R., & Avery, L. W. (1983). U.S. Army Test and Evaluation Command test operations procedures. (Part I-Test Procedures, Part II-HEDGE, TOP 1-2-610). Aberdeen, MD: U.S. Army Test and Evaluation Command.

Pew, R. W., Baron, S., Feehrer, C. E., & Miller, D. C. (1977). Critical review and analysis of performance models applicable to man-machine systems evaluation (BBN Report No. 3446). Bolling AFB, DC: Air Force Office of Scientific Research.

Rose, A. M. (1978). An information processing approach to performance assessment (Rep. No. AIR 58500-11/78-FR). Washington, DC: American Institutes for Research.

Schendel, J. D., Shields, J. L., & Katz, M. S. (1978). Retention of motor skills: Review (Technical Paper 313). Alexandria, VA: Army Research Institute. AD A061-338

Schoenfeldt, L. F. (1982). Intra-individual variation and human performance. In M. D. Dunnette and E. A. Fleishman (Eds.), Human performance and productivity: Human capability assessment. (Vol. 1, pp. 107-135). Hillsdale, NJ: Lawrence Erlbaum.

Simon, C. W. (1976). Analysis of human factors research engineering experiments (TR-CWS-02-76). Westlake Village, CA: Canyon Research Group.

Sims, W., & Hiatt, C. M. (1981). Validation of the Armed Services Vocational Aptitude Battery (ASVAB) forms 6 and 7 with applications to ASVAB forms 8, 9, and 10. Alexandria, VA: Center for Naval Analyses. (NTIS No. AD A110 025)

Smith, R. S., & Kennedy, R. S. (1976). Predictor displays: A human engineering technology in search of a manual control program (PMTC 76-05). Point Mugu, CA: Pacific Missile Test Center.

Snow, R. E. (1980). Aptitudes and instructional methods: Research on individual differences in learning-related processes (Final Report 1975-1979, Aptitude Research Project), Stanford, CA: School of Education, Stanford University.

Swain, A. D., & Wohl, J. G. (1961). Factors affecting degree of automation in test and checkout equipment. Stamford, CT: Dunlap and Associates.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. Psychological Review, 68, 301-340.

Tice, J. (1986). Army program integrates personnel training and weapons design. Defense News, January 27, p. 6.

U.S. General Accounting Office. (1985). The Army can better integrate manpower, personnel, and training into the weapon systems acquisition process. Report to the Secretary of the Army, GAO/NSIAD-85-154.

Wallace, J. R. (1982). The Gideon criterion: The effects of selection criteria on soldier capabilities and battle results (USAREC Research Memorandum 82-1). Fort Sheridan, IL: U.S. Army Recruiting Command.

Westra, D. P., & Lintern, G. (1985). Simulator design features for helicopter landing on small ships. I. Performance Study I (NAVTRASYSCEN 81-C-0105-13). Orlando, FL: Naval Training Equipment Center.

Westra, D. P., Simon, C. W., Collyer, S. C., & Chambers, W. S. (1982). Simulator design features for carrier landing: I. Performance experiments (NAVTRAEQUIPCEN 78-C-0060-7). Orlando, FL: Naval Training Equipment Center.

Wightman, D. C., & Lintern, G. (1985). Part-task training for tracking and manual control. Human Factors, 27(8), 267-283.

Winer, B. J. (1971). Statistical principles in experimental design (2nd Ed.). New York: McGraw-Hill.

Woodson, W. (1955). 100 rules for improving man-machine effectiveness. San Diego, CA: TeleTech Electronics Industries.

Yalow, E. (1980). Individual differences in learning from verbal and figural materials (Technical Report No. 12, Aptitude Research Project). Stanford, CA: Stanford University.

ADDITIONAL READINGS

Armstrong, H. G. (1943). Principles and practice of aviation medicine. Baltimore, MD: The Williams & Wilkins Company.

Army Ballistic Missile Agency. (1959). PERSHING development specification for weapon system human factor design criteria (ABMA-XPD-844). Redstone Arsenal, AL.

Army Ballistic Missile Agency (1961). Weapon systems human factors engineering criteria (ABMA-STD-434). Redstone Arsenal, AL.

Boff, K. R. & Lincoln, J. E. (1986) Human engineering data compendium for designers. Wright-Patterson AFB, OH: Air Force Armstrong Aerospace Medical Research Laboratory.

Carter, R. C., & Biersner, R. J. (1985). Validation of job attribute requirements derived from the Position Analysis Questionnaire using military aptitude test scores. Unpublished manuscript. New Orleans, LA: Naval Biodynamics Laboratory.

Chaikin, G. (1978). Human engineering design criteria - the value of obsolete standards and guides (Technical Note 13-78). Aberdeen Proving Ground, MD: U.S. Army Human Engineering Laboratory.

Chaikin, G., & Chaillet, R. F. (1965). Engineering practice study-human engineering. Department of Defense Standardization Program Project 1410-0016, U.S. Army Missile Command Report No. RC-S-65-1, Redstone Arsenal, AL.

Chapanis, A., Garner, W. R., & Morgan, C. T. (1949). Applied experimental psychology: Human factors in engineering design. New York: Wiley.

Committee on Undersea Warfare. (1949). A survey report on human factors in undersea warfare (Rep. No. N7-ORR-291). Washington, DC: National Research Council.

Department of Defense. (1966). Military Standard H-81444(AS). Military specifications for system analysis data (MIL-H-8144 AS). Washington, DC.

Department of Defense. (1970). Military Standard H-46855. Human engineering requirements for military systems, equipment and facilities (MIL-H- 46855). Washington, DC.

Department of the Air Force (1959). Human engineering criteria for aircraft, missile, and space systems ground support equipment. MIL-STD-803, Wright-Patterson Air Force Base, OH.

Department of the Air Force (1964). Human engineering design criteria for aerospace systems and equipment: Part 1. Aerospace system ground equipment. MIL-STD-803A1, Washington, DC.

Loo, P. (1978). Individual differences and the perception of traffic signs. Human Factors, 20(1), 65-74.

McFarland, R. A. (1953). Human factors in air transportation: Occupational health and safety. New York, NY: McGraw-Hill.

Promisel, D. M., Hartel, C. R., Kaplan, J. D., Marcus, A., & Whittenberg, J. A. (1985). Reverse engineering: Human factors, manpower, personnel, and training in the weapon system acquisition process (Technical Rep. 659). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. AD A157-693

Reid, G. B. (1982). Subjective workload assessment: A conjoint scaling approach. Proceedings of the 1982 Annual Scientific Meeting: Aerospace Medical Association (p. 154). Washington, DC: Aerospace Medical Association.

Shingledecker, C. A. (1982). Performance evaluation of the embedded secondary task technique. Proceedings of the 1982 Annual Scientific Meeting, Aerospace Medical Association (pp. 151-152). Washington, DC: Aerospace Medical Association.

Taylor, F. V. (1963). Human engineering and psychology. In S. Koch (Ed.), Psychology: A study of science (Vol. 5, pp. 831-907). New York: McGraw-Hill.

U.S. Air Force (1958). Human engineering design standards for missile system equipment (AFBM Exhibit 57-3A). Air Force Ballistic Missile Division, Inglewood, CA.

Vaughan, G. M., & Corballis, M. C. (1969). Beyond tests of significance: Estimating strength of effects in scheduled ANOVA designs. Psychological Bulletin, 72, 204-213.

Wickens, C. D. (1984). Engineering psychology and human performance. Columbus, OH: Charles E. Merrill.

END

DATE

FILMED

2-89